RECOGNITION AND ASSEMBLY USING PROTEIN "BUILDING BLOCKS"

Kevin P. McGrath*, David L. Kaplan

Biotechnology Division, U.S. Army Natick Research, Development & Engineering Center, Natick, Massachusetts USA 01760-5020

Abstract: A new approach to materials design is presented, utilizing specific recognition and assembly of proteins at the molecular level. The approach exploits the control over polymer chain microstructure afforded by biosynthesis to produce protein-based materials with precisely defined physical properties. Incorporated into these materials are recognition elements that stringently control the placement and organization of each chain within higher order superstructures. The proteins, designated Recognin A2 through Recognin E2, are recombinant polypeptides designed de novo from both natural consensus sequences and an appreciation of the physical principles governing biological recognition. The synthesis and characterization of the protein recognition elements is briefly described and initial studies on self-assembly-recognition patterns using surface plasmon resonance and circular dichroism are presented. A subset of these materials are programmed to spontaneously assembly into complex, multicomponent structures and represent a first step in a rational approach to nanometer-scale structural design.

INTRODUCTION

Efficient engineering at the molecular scale requires individual molecules to recognize specific counterparts in multicomponent systems and spontaneous organization into well-defined molecular complexes of hundreds of thousands of molecules. Traditional materials science rarely controls organization down to the molecular level and the ordering of individual molecules is statistical in nature. Considerable post-assembly processing is often required to achieve desired physical properties. In order to successfully fabricate materials at the nanometer-scale we must rely on the molecules to process themselves into useful assemblies. This can be accomplished by incorporating into each component the ability to spontaneously recognize where it belongs in a larger framework and to incorporate itself into the final assembly in a controlled fashion.

With this need in mind, natural systems can serve as a guide due to the ubiquitous role of molecular recognition in biological processes. Recognition patterns based on shape, charge placement, hydrophobicity, and other interactions are common in many natural processes. The molecular-level tailorability and controlled synthesis of protein-polymers afford unprecedented opportunities in the study of recognition processes. One class of proteins with well defined

 CCC 1022-1360/94/$ 04.00

primary and secondary structures are the transcription factors (e.g., mammalian - C/EBP, Jun, Fos; yeast - GCN4, YAP1) which form a heterodimeric coil-coiled motif to perform their functions as regulators of DNA activity. They are all dimeric in their functional form. Two aspects of recognition are encoded in the primary sequence of these proteins, congener recognition through "leucine zippers" to form a hydrophobic core surrounded by charge interactions, and DNA-protein recognition to control protein-DNA binding and interactions during transcription. We have chosen this coil-coiled structural motif as a starting point to develop an understanding of molecular recognition as a route to more complex architectures through self-assembly of proteins.

A number of research groups are exploring the functions of these types or proteins and these efforts have led to a better understanding of recognition in terms of the leucine interactions, the role of charged residues in neighboring positions, and the parallel vs. antiparallel arrangement of congener or dimer pairs (Refs. 1-5). Our approach has been to design a series of genes encoding proteins based on this motif wherein the specific placement of the charged residues is defined and controlled. This provides an opportunity to use a family of protein building blocks to study protein-protein recognition and self-assembly and to identify the role of charge placement in this process. We will briefly describe the design of the genetic elements encoding these proteins and the expression and purification of these building blocks. We will focus on initial studies on protein-protein recognition carried out with these recombinant materials.

RESULTS AND DISCUSSION

Oligonucleotides were synthesized on a Milligen/Biosearch Cyclone Plus DNA Synthesizer using the phosphoramidite chemistry of McBride and Caruthers (Ref. 6). The synthetic DNA fragments were inserted into pUC18 and used to transform Escherichia coli strain NM522 for amplification (Ref. 7). Color selection on plates containing isopropyl thiogalactoside (IPTG) and 5-bromo-4-chloro-3-indolyl-ß-galactopyranoside (X-Gal) was used to identify positive recombinants. Plasmids containing inserts were sequenced to validate orientation. A synthetic 57 base pair linker fragment was constructured and inserted into pUC18. The isolated fragments were combined to generate various 126 base pair DNA recognition elements. These were constructed in situ by combining and ligating the desired fragments in the presence of linearized vector. The recombinant plasmids were verified by double-stranded plasmid sequencing.

The appropropriate recognition elements were excised from the recombinant pUC-BstEII, gel purified, inserted into pQE-9 (Qiagen, Inc., Chatsworth, CA), and used to transform E. coli. Restriction digests were used to check orientation of inserts and the correct inserts were used to transform E. coli SG13009 PREP4 (Qiagen, Inc.). The protein recognition elements (termed Recognins) were expressed in tryptone, yeast extract, NaCl medium with appropriate

antibiotics. Protein production was induced with IPTG. Proteins were purified using immobilized metal affinity chromatography and yields ranged from 70 to 140 mg/liter.

The purified protein recognition elements (Recognins A2, B2, C2, D2, E2) were characterized by amino acid composition, N-terminal sequencing, capillary electrophoresis, and matrix-assisted laser desorption mass spectrometry. Purities in excess of 95% were found for all proteins and molecular weights were from 13,105 to 13,127 for some of the proteins based on laser desorption characterization.

With the family of Recognin proteins available through the processes described, precise control over the specificity of assembly and recognition was explored. Based on the design of the Recognin proteins, regularly spaced hydrophobic residues (leucines) are generated within a heptad repeat $(\underline{a}\text{-}b\text{-}c\text{-}\underline{d}\text{-}e\text{-}f\text{-}g)_n$. Within this repeat, hydrophobic residues are found at the a and d positions and polar and charged amino acid residues occupy the e and g positions. The alpha-helices wrap around each other with a slight superhelical twist having a periodicity of about 140Å. An isolated coiled coil is about 10Å thick. Dimerization is driven by the sequestering of hydrophobic residues in the interior of the coiled-coil, with carefully positioned charged residues providing the specificity in partners. Site-directed mutagenesis of these recognition regions has demonstrated that while the driving force for dimerization is hydrophobic, the dominant factor in preferential heterodimer formation was electrostatic interactions occuring between sidechains at the e and g positions in the coiled-coil (Ref. 8). The degree of interaction in the Fos-Jun congener pair has been studied using short synthetic peptides identical in sequence to the leucine zipper portions (Ref. 1). Using HPLC it was shown that heterodimer preference was at least 100X higher than homodimer preference. They also confirmed prior studies (Ref. 2) that the two alpha-helices align in parallel, vs. the antiparallel arrangement reported originally (Ref. 9). This conclusion has been further supported with X-ray diffraction data from crystals of the homodimer GCN4 (Ref. 3).

The library of protein sequences generated in the present study are encompassed by the generic primary sequences shown in Fig. 1. The identity and position of the leucines and charged groups are in agreement with natural leucine zipper proteins, however a number of natural proteins that contain repeated leucine heptads do not form coiled coils. Therefore, an algorithm was used to identify coiled-coils from primary sequence data to refine the protein sequences to select those with a high probability of coiled coil formation (Ref. 10). Since we used a mixed site approach for the first base of the codons for amino acids at the e and g positions in the gene construction step we generated a library of DNA sequences encoding all 4094 (2^{12}) possible recognition sequences. The pattern of charged residues (glutamates and lysines) was designed to discourage homodimerization and to promote a parallel in-register orientation of the two helices. With the crystallographic data from Kim and coworkers (Ref. 3) the protein building blocks were chosen as shown in Fig. 1.

Generic Sequence:
(IGDL[E/K]N[E/K]VAQL[E/K]R[E/K]VRSL[E/K]D[E/K]
AAEL[E/K]Q[E/K]VSRL[E/K]N[E/K]IEDL[E/K]A[E/K])$_n$

Recognin Charge Sequences:

A2:	EEEE	EEEK	KEKE	EEEE	EEEK	KEKE
B2:	KKKK	KKKE	EEKK	KKKK	KKKE	EEKK
C2:	KKEE	KKKE	EEKK	KKEE	KKKE	EEKK
D2	EEEE	KKKE	EEKK	EEEE	KKKE	EEKK
E2	EEEE	EEKE	EEKK	EEEE	EEKE	EEKK

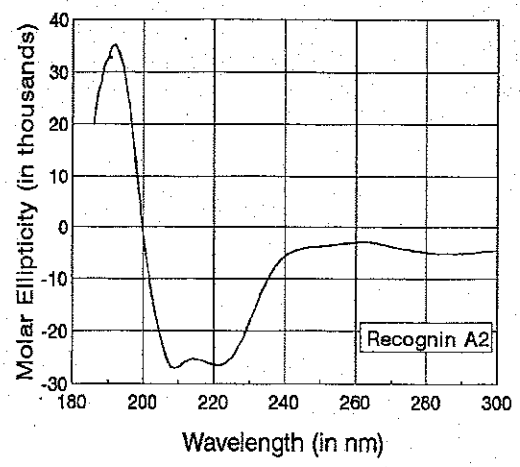
Fig. 1. Library of Recognin sequences. Top: Generic Recognin coiled-coil sequence with leucines underlined. Bottom: Charge patterns for Recognins A2 through E2 where only the e and g residues are shown.

Crystallographic data of the GCN4 homodimer indicates that electrostatic bonding occurs between the e residue of the ith heptad and the g residue of the ith-1 heptad. Using this information, the Recognin sequences were designed such that A2 is the target sequence and B2 is a perfect electrostatic fit with A2. Recognins C2, D2, and E2 were designed to be progressively poorer fits with the charge pattern in A2. The specificity, strength and stability of the recognition events were studied by turbidity, surface plasmon resonance, and circular dichroism.

Turbimetric measurements were taken on Beckman DU-70 UV/Vis spectrophotometer at 400 nm using A2 and B2. The results showed a rapid increase in turbidity upon mixing equal concentrations of the two Recognins. This result was attributed to specific heterodimer formation since no similar increase in turbidity occurs with the individual Recognins. Thus it appears that specific heterodimerization is favored due to the increased stability of the A2-B2 complex when compared to A2-A2 or B2-B2. Gel electrophoresis analysis of the precipitate from the heterodimer reaction precipitate indicates that it is a 1:1 complex of A2 and B2.

The analysis of the Recognins in solution by circular dichroism on an AVIV 60DS spectrophotometer is presented in Fig. 2. An example of the absorption curve for A2 in 0.1M KF is presented along with the calculated helix content for all five Recognins. The high helix content confirms the predicted secondary structure for these proteins based on their design. Melting transitions for helix-random coil conformations for various pairings of the proteins are presented in Tab. 1. These values were determined in 6M guanidine hydrochloride, 0.1M PO_4, pH 8.0, at 222 nm over a temperature scanning range of 0°C to 70°C. The ΔT_m is higher for

A2-B2 than for A2-C2, indicative of an increased stability between A2-B2. The ΔT_m is calculated as the difference between the transition for a given pairing minus the average transition for the individual component pairs.



Helix Content

A2	80.6%
B2	74.5%
C2	83.5%
D2	73.8%
E2	65.6%

Fig. 2. Circular dichroism data for A2 and the helix content calculated for the five Recognins.

Interaction	K_a	Selectivity	ΔT_m
A2-B2	$3.14X10^6$	3.26	+3.8°C
A2-C2	$1.45X10^6$	1.50	+2.3°C
A2-A2	$9.64X10^5$	1.00	---

Tab. 1. Solution interactions between Recognins in 6M guanidine hydrochloride.

Surface plasmon resonance studies between A2 and B2 or C2 were conducted on a Biacore Biospecific Interaction Analysis System (Pharmacia Biosensor, Upsala, Sweden). Kinetic analyses were performed where one of the Recognins was immobilized on a thin gold film. This provides an opportunity to measure the rates of association and dissociation at the picogram level between the complexes based on small changes in refractive index related to the binding event in the vicinity of the metal surface. Fig. 3 illustrates the relative differences in interactions among A2-B2 and A2-C2. The data for association constants and selectivity are

shown in Tab. 1. The selectivity of A2 for B2 is higher than the selectivity of A2 for C2, and the equilibrium response is achieved sooner. The rates of dissociation observed in the later portions of the sensorgram are not very different.

We have shown through genetic engineering that a series of proteins (Recognins) can be synthesized, expressed and purified from a bacterial host. These proteins, designed for specific recognition events, are capable of self-assembly into coiled-coil motifs as anticipated based on their design. The degree of association can be correlated with the placement of charged residues

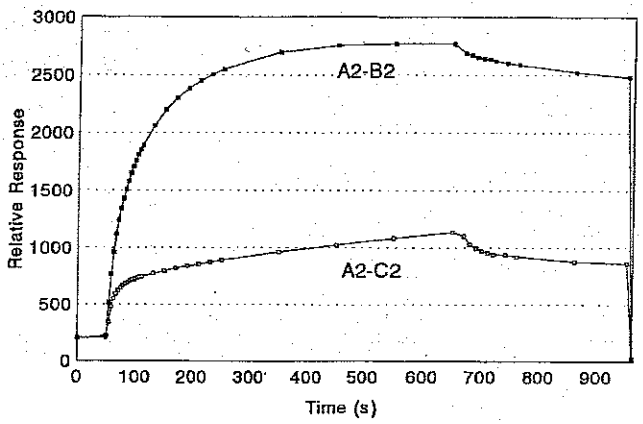


Fig.3. Surface plasmon data for A2-B2 and A2-C2 interactions. The initial increase in refractive index is due to solvent exchange.

in the heptad repeats such that initial evidence both in the immobilized and in solution states indicate relative differences in affinity among the different protein designs. This degree of control, if sufficiently understood, serves as a starting point for the controlled self-assembly of more complex materials with precisely controlled architectures. With appropriate head-to-tail and side-to-side recognition elements built into the encoded proteins the increase in architectural complexity can be anticipated. In the long run, the potential high level of control over molecular architecture through these materials shoud permit the precise tailoring of macroscopic/functional properties such as mechanical performance, thermal performance, barrier performance, and other physical, chemical and biological functions and responses.

ACKNOWLEDGMENT

Carla DiGirolamo for help with the genetics and protein expression work and to John Walker for help with the circular dichroism studies.

REFERENCES

(1) E. K. O'Shea, R. Rutkowski, W. F. Stafford, P. S. Kim, Science 245, 646 (1989)
(2) E. K. O'Shea, R. Rutkowski, P. S. Kim, Science 243, 538 (1989)
(3) E. K. O'Shea, J. D. Klemm, P. S. Kim, T. Alber, Science 254, 539 (1991)
(4) R. S. Hodges, N. E. Zhou, C. M. Kay, P. D. Semchuk, Peptide Res. 3, 123 (1990)
(5) N. Z. Zhou, B.-Y. Zhu, C. M. Kay, R. S. Hodges, Biopolymer 32, 419 (1992)
(6) L. J. McBride, M. H. Caruthers, Tetrahedron Lett. 24, 245 (1983)
(7) J. Sambrook, E. F. Frisch, T. Maniatis, Molecular Cloning, A Laboratory Manual, 2nd Ed., Cold Spring Harbor (1989)
(8) M. Scherman, J. B. Hunter, G. Hennig, R. Muller, Nucleic Acid Res. 19(4), 739 (1991)
(9) W. H. Landschulz, P. F. Johnson, S. L. McKnight, Science 240, 1759 (1988)
(10) A. Lupas, M. Van Dyke, J. Stock, Science 252, 1162 (1991)